Package 'CIMLR'

October 15, 2018

Version 1.0.0

Date 2018-10-14

Title Cancer Integration via Multikernel Learning (CIMLR)

Maintainer Luca De Sano <luca.desano@gmail.com>

Depends R (>= 3.5),

Imports parallel, Matrix, stats, methods, Rcpp, pracma, RcppAnnoy, RSpectra

Suggests BiocGenerics, BiocStyle, testthat, knitr, igraph

Description Outcomes for cancer patients vary greatly even within the same tumor type, and characterization of molecular subtypes of cancer holds important promise for improving prognosis and personalized treatment. This promise has motivated recent efforts to produce large amounts of multidimensional genomic ('multi-omic') data, but current algorithms still face challenges in the integrated analysis of such data. In this package we provide the implementation of Cancer Integration via Multikernel Learning (CIMLR), a new cancer subtyping method that integrates multi-omic data to reveal molecular subtypes of cancer.

Encoding UTF-8

LazyData TRUE

License file LICENSE

URL https://github.com/danro9685/CIMLR

BugReports https://github.com/danro9685/CIMLR

biocViews Clustering, CancerData

RoxygenNote 6.1.0

LinkingTo Rcpp

NeedsCompilation yes

VignetteBuilder knitr

Author Daniele Ramazzotti [aut, cre],
Avantika Lal [aut],
Bo Wang [ctb],
Luca De Sano [aut],
Serafim Batzoglou [ctb],
Arend Sidow [ctb]

R topics documented:

CIMLR *CIMLR*

Description

perform the CIMLR clustering algorithm

Usage

```
CIMLR(X, c, no.dim = NA, k = 10, cores.ratio = 1)
```

Arguments

X	a list of multi-omic data each of which is an (m x n) data matrix of measurements of cancer patients
c	number of clusters to be estimated over X
no.dim	number of dimensions
k	tuning parameter
cores.ratio	ratio of the number of cores to be used when computing the multi-kernel

Value

clusters the patients based on CIMLR and their similarities

list of 8 elements describing the clusters obtained by CIMLR, of which y are the resulting clusters: y = results of k-means clusterings, S = similarities computed by CIMLR, F = results from network diffiusion, ydata = data referring the the results by k-means, alphaK = clustering coefficients, execution.time = execution time of the present run, converge = iterative convergence values by T-SNE, LF = parameters of the clustering

Examples

```
CIMLR(X = GliomasReduced$in_X, c = 3, cores.ratio = 0)
```

CIMLR_Estimate_Number_of_Clusters
CIMLR Estimate Number of Clusters

Description

estimate the number of clusters by means of two huristics as discussed in the CIMLR paper

Usage

```
CIMLR_Estimate_Number_of_Clusters(all_data, NUMC = 2:5,
  cores.ratio = 1)
```

Arguments

all_data	is a list of multi-omic data each of which is an (m x n) data matrix of measurements of cancer patients
NUMC	vector of number of clusters to be considered
cores.ratio	ratio of the number of cores to be used when computing the multi-kernel

Value

a list of 2 elements: K1 and K2 with an estimation of the best clusters (the lower values the better) as discussed in the original paper of SIMLR

Examples

```
CIMLR_Estimate_Number_of_Clusters(GliomasReduced$in_X,
   NUMC = 2:5,
   cores.ratio = 0)
```

CIMLR_Feature_Ranking *CIMLR Feature Ranking*

Description

perform the CIMLR feature ranking algorithm. This takes as input the combination of the original input data appended by rows and the corresponding similarity matrix computed by CIMLR

Usage

```
CIMLR_Feature_Ranking(A, X)
```

Arguments

A	an (n x n) similarity matrix by CIMLR
X	a set of multi-omic data inputs each of which is an (m x n) data matrix of measurements of cancer patients appended by rows

Value

a list of 2 elements: pvalues and ranking ordering over the n covariates as estimated by the method

Examples

```
cimlr = CIMLR(X = GliomasReduced$in_X, c = 3, cores.ratio = 0)
input_data = rbind(GliomasReduced$in_X$point_mutations,GliomasReduced$in_X$copy_numbers,
                GliomasReduced$in_X$methylations,GliomasReduced$in_X$expression_values)
CIMLR_Feature_Ranking(A = cimlr$S, X = input_data)
```

GliomasReduced *test dataset for CIMLR*

Description

example dataset to test CIMLR. This is a reduced version of the dataset from the work by The Cancer Genome Atlas Research Network.

Usage

```
data(GliomasReduced)
```

Format

multi-omic data of cancer patients

Value

list of 1 element: in_X = input dataset as a list of 4 (reduced) multi-omic data each of which is an (m x n) measurements of cancer patients

Source

Cancer Genome Atlas Research Network. "Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas." New England Journal of Medicine 372.26 (2015): 2481-2498.

Index